

Protein Structure Analysis

<http://binf.gmu.edu/vaisman/binf731/>

Iosif Vaisman

2009

Secondary structure characterization
Secondary structure assignment
Secondary structure prediction
Protein structure classification

Secondary Structure Prediction

Three-state model: helix, strand, coil

Given a protein sequence:

- NWVLSTAADMQGVVTDGMASGLDKD . . .

Predict a secondary structure sequence:

- LLEEEELLLLHHHHHHHHHLLHHHL . . .

Methods:

- statistical
- stereochemical

Accuracy: 50-85%

Statistical Methods

Residue conformational preferences:

Glu, Ala, Leu, Met, Gln, Lys, Arg - helix
Val, Ile, Tyr, Cys, Trp, Phe, Thr - strand
Gly, Asn, Pro, Ser, Asp - turn

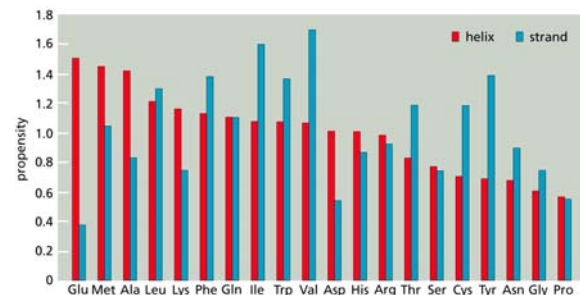
Chou-Fasman algorithm:

Identification of helix and sheet "nuclei"
Propagation until termination criteria met

Chou-Fasman Parameters

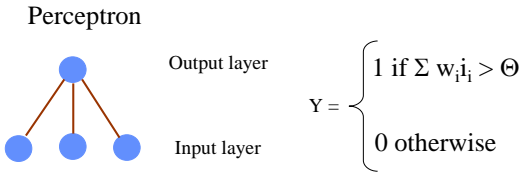
Name	P(a)	P(b)	P(turn)	f(i)	f(i+1)	f(i+2)	f(i+3)
Alanine	142	83	66	0.06	0.076	0.035	0.058
Arginine	98	93	95	0.070	0.106	0.099	0.085
Aspartic Acid	101	54	146	0.147	0.110	0.179	0.081
Asparagine	67	89	156	0.161	0.083	0.191	0.091
Cysteine	70	119	119	0.149	0.050	0.117	0.128
Glutamic Acid	151	37	74	0.056	0.060	0.077	0.064
Glutamine	111	110	98	0.074	0.098	0.037	0.098
Glycine	57	75	156	0.102	0.085	0.190	0.152
Histidine	100	87	95	0.140	0.047	0.093	0.054
Isoleucine	108	160	47	0.043	0.034	0.013	0.056
Leucine	121	130	59	0.061	0.025	0.036	0.070
Lysine	114	74	101	0.055	0.115	0.072	0.095
Methionine	145	105	60	0.068	0.082	0.014	0.055
Phenylalanine	113	138	60	0.059	0.041	0.065	0.065
Proline	57	55	152	0.102	0.301	0.034	0.068
Serine	77	75	143	0.120	0.139	0.125	0.106
Threonine	83	119	96	0.086	0.108	0.065	0.079
Tryptophan	108	137	96	0.077	0.013	0.064	0.167
Tyrosine	69	147	114	0.082	0.065	0.114	0.125
Valine	106	170	50	0.062	0.048	0.028	0.053

Chou-Fasman Parameters



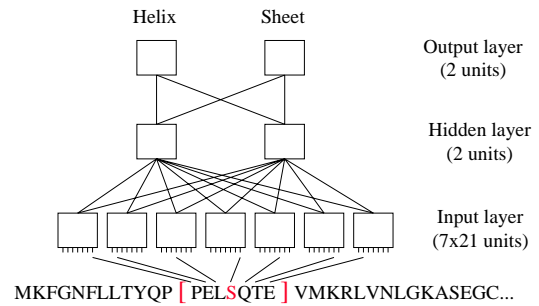
Adopted from Zvelebil, Baum, 2008

Neural Networks



Learning process: $\Delta w_i = (T_p - Y_p) i_{pi}$

Neural Networks Methods



Stereochemical Methods

Patterns of hydrophobic and hydrophilic residues in secondary structure elements:

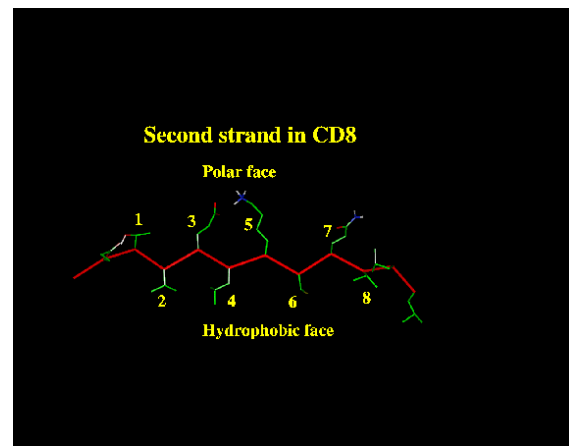
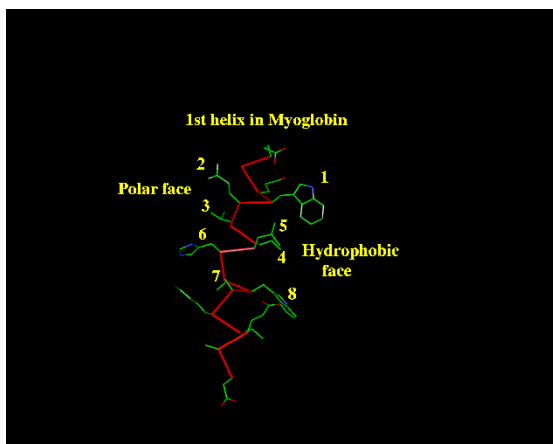
- segregation of hydrophobic and hydrophilic residues
- hydrophobic residues in the positions 1-2-5 and 1-4-5
- oppositely charged polar residues in the positions 1-5 and 1-4 (e.g. Glu (i), Lys (i+4))

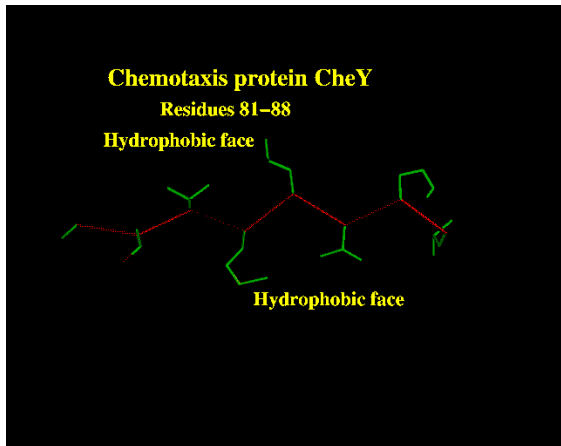
Definitions of hydrophobic and hydrophilic residues (hydrophobicity scales) are ambiguous

Stereochemical Methods

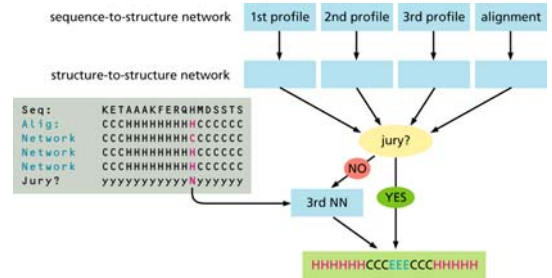
Hydropathic correlations in helices and sheets

		F-F	F-L	L-F	L-L
α	$i, i+2$	-	+	+	-
	$i, i+3$	+	-	-	+
	$i, i+4$	+	-	-	+
	$i, i+5$	-	+	+	-
β	$i, i+1$	-	+	+	-
	$i, i+2$	+	-	-	+
	$i, i+3$	-	+	+	-

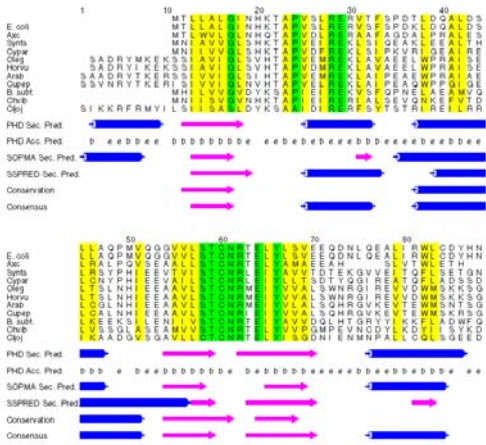




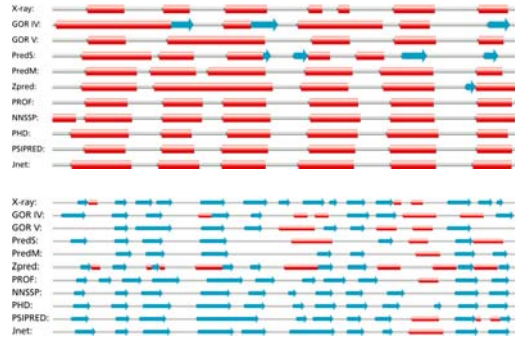
Jnet Algorithm



Adopted from Zvelebil, Baum, 2008

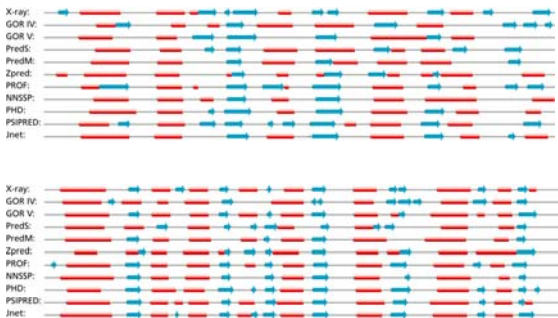


Accuracy of prediction



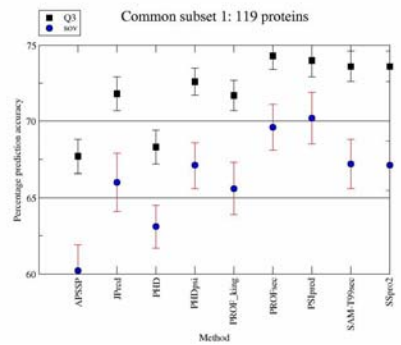
Adopted from Zvelebil, Baum, 2008

Accuracy of prediction



Adopted from Zvelebil, Baum, 2008

Accuracy of prediction



EVA (<http://cubic.bioc.columbia.edu/eva/>)

Accuracy of Prediction

$$Q_3 = \frac{PH + PE + PC}{N}$$

$$W = \log \frac{TP \times TN}{FP \times FN}$$

Range: 50-85%